DataCube

Email: info@apperception.co.uk
Tel: 0845 644 3479
Web: www.apperception.co.uk/datacube

Conceptual Categorisation

- **Makes content classification straightforward**
- **Analyses a document's actual meaning rather than just in dividual words or strings**
- **Offers Taxonomy/Schema Options:**
  - **Existing**
  - **Bespoke**
  - **Auto generated**
- **Incredibly accurate and lightning fast**
- **Essential preparation for migrating data to EDRMS or the Cloud**

Organisational Requirements

Many organisations experience significant difficulties with regard content classification or categorisation. Often, months are wasted developing what is believed to be an appropriate schema only for it to be discovered that, on applying it, the vast majority of documents have been categorised into so few of the selected headings that it is unusable.

DataCube Solution

DataCube eliminates the time-consuming burden of content classification or categorization and enables an organisation's staff to quickly find the information they need and then use it to inform decisions or actions. Unlike traditional search engines, DataCube technology can automatically populate categories with documents based on the concepts found within those documents.

Features

- Works across multiple EDRMS systems and unstructured data sets
- Allows access to and analysis of whole organisation data sets, not just departmental silos
- Automatically achieves what is impossible in a manual exercise
- Can feed the results into other applications such as SQL Server, SharePoint or Dynamics, or a document handling system
- Create communities of subject matter or automatically cluster content and resources into related groups.

Email: info@apperception.co.uk
Tel: 0845 644 3479
Web: www.apperception.co.uk/datacube

DataCube

How Does it Work?

Concept-based Categorization is a remarkably simple two-stage process that mimics how people think and classify documents:

The first step is to identify the categories into which the documents are desired to be sorted or classified. Categories are conceptual in nature, can be tightly focused or very broad and can be named however an organisation wishes. There is no limit to the number of categories into which documents can be classified and the classification hierarchy can be an existing schema, a bespoke one or one automatically generated by the DataCube itself based on its analysis of the conceptual content

Once that is done, the next stage is to identify example documents (exemplars) for each category. An examplar can be the whole or part of a document that is representative of the category. To help craft effective sample sets, DataCube provides a "self-test" that compares all the example documents in a given category and points out any that are not a good "fit" within that category. For Local Authorities, we have already developed an implementation ready schema that includes over 18,000 exemplars that fully populate the 750 categories in the Local Government Classification Scheme (LGCS).

With the completion of those two steps, concept-based categorization is totally automatic. Users can control how generalized or specific the "fit" into a category should be, and even whether or not a document should fall into multiple categories if it contains information that is relevant to more than one category. Users can also set thresholds that will determine whether DataCube categorizes a document into a pre-identified category or drops it into an "uncategorized" folder for later review. As with all its functions, DataCube provides a numeric score to tell users just how well each document fits into the resultant categories.

Organisational Benefits

- Provides access to all of an organisation's data regardless of the number of business systems- leaving original document in situ
- Far better integration of information allowing better decision making
- Offers greater efficiencies in business processes such as Freedom of Information request handling
- Enables a real "one stop shop" approach to customer care